



To you who are reading, hello!

My name is Beatrice Occhiena and I am a student of Artificial Intelligence.

Since I study so many different subjects on my own, I tend to rely a lot on all the resources I find on the web [🌐](#). Whenever I do some research online to learn something new, I always feel very lucky, because I find so much free material carefully crafted by other people from all over the world. So I too decided to make my own small contribution ✨.

Over the months I have put a lot of care into the creation of these notes ↗. And although I know that they may be incomplete and imperfect in some places, I would like to make them available to you — in the hope that they may help even one student in the exam preparation.

I firmly believe in collaboration and mutual support between students ❤️. Furthermore, in the future I would love to be involved in research and education, so this is a first opportunity for me to showcase my work.

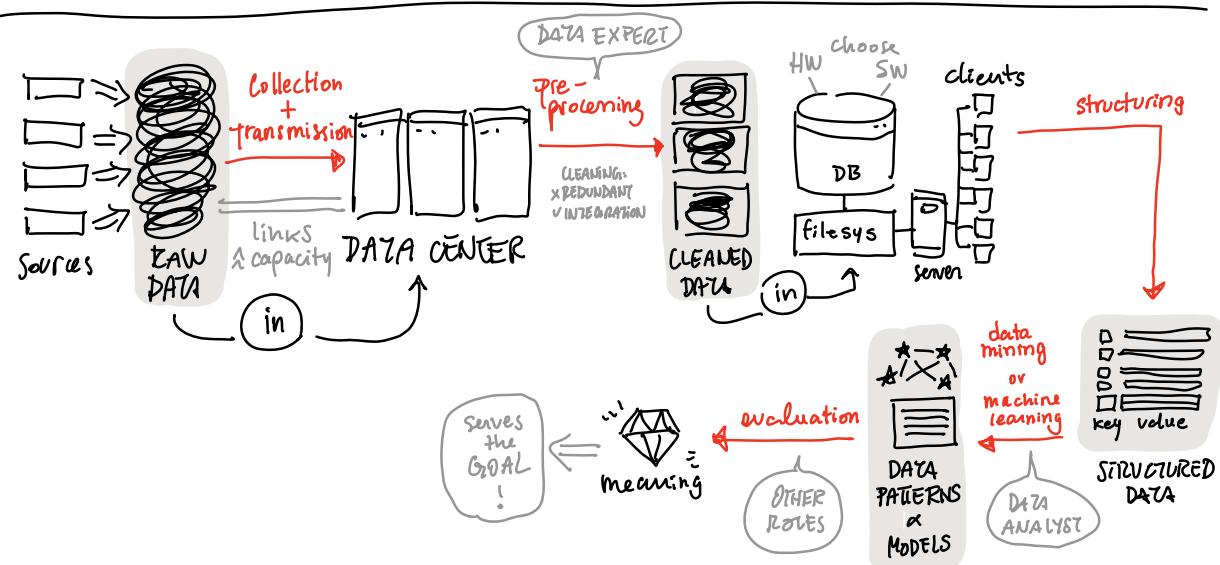
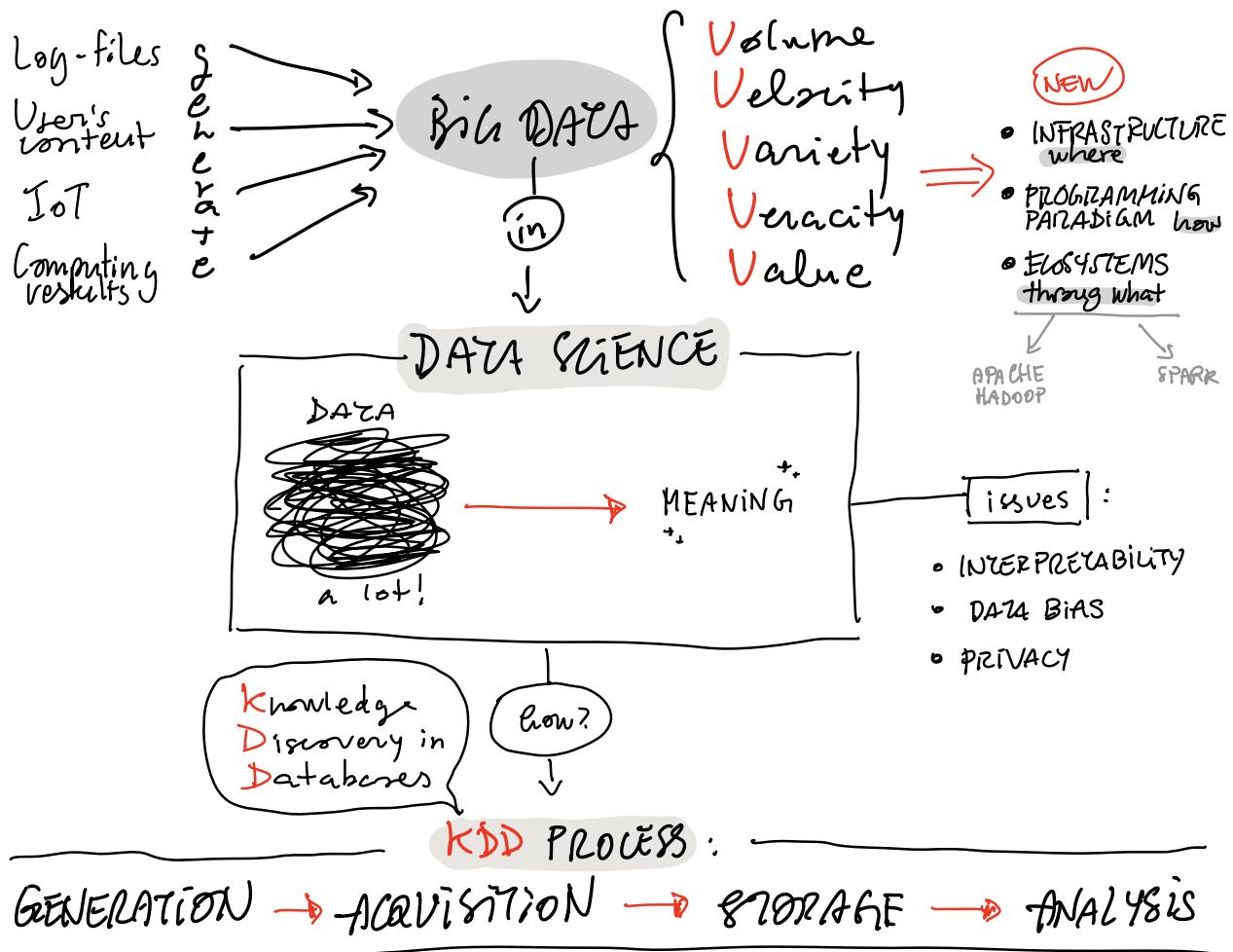
That said, I wish you all the best for this exam! Take these words as a small gesture of encouragement. Even though we don't know each other, I still want to cheer you on! Do your best and don't forget to have as much fun as possible studying this subject 🎉.

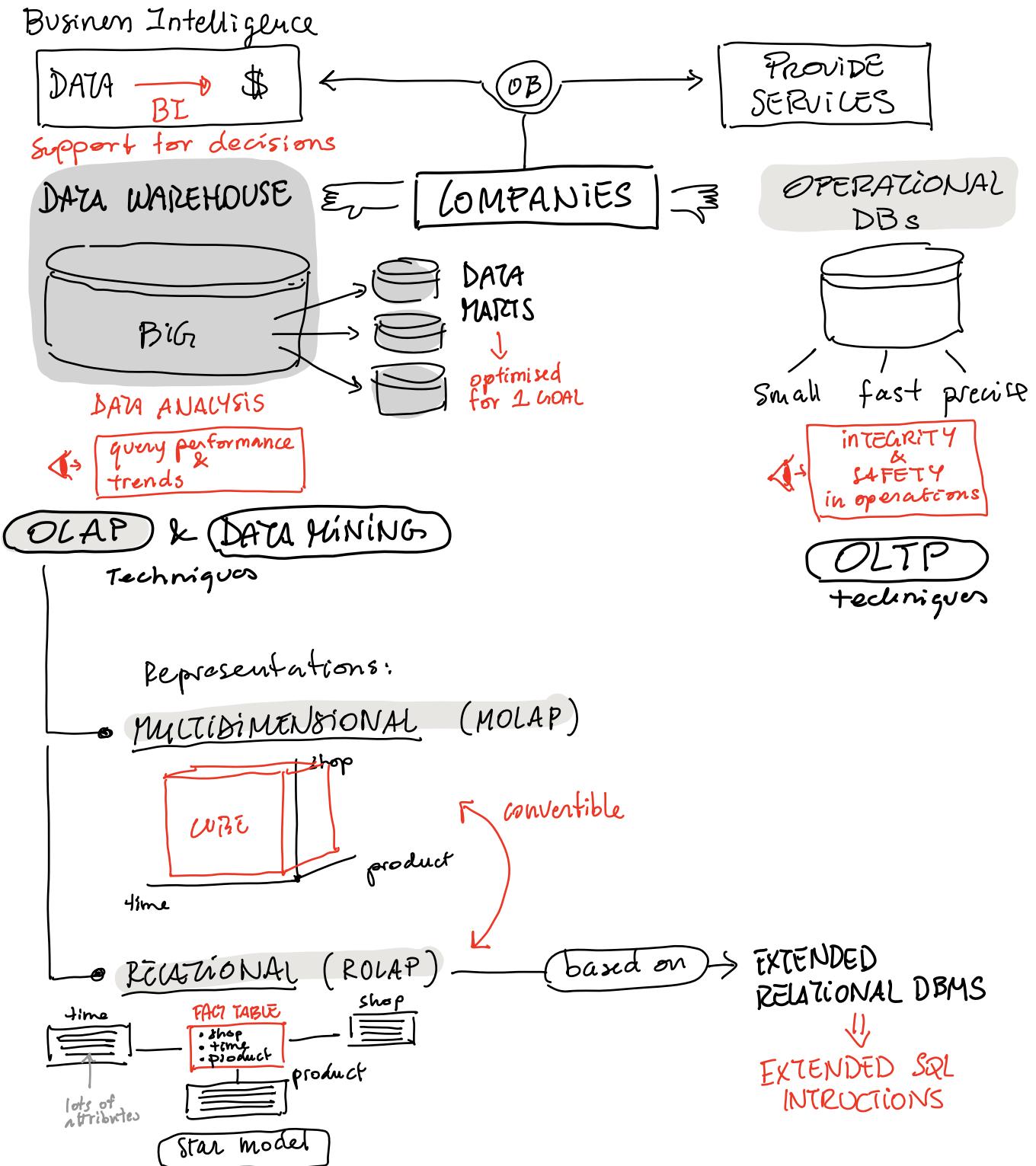
— Contact me! —

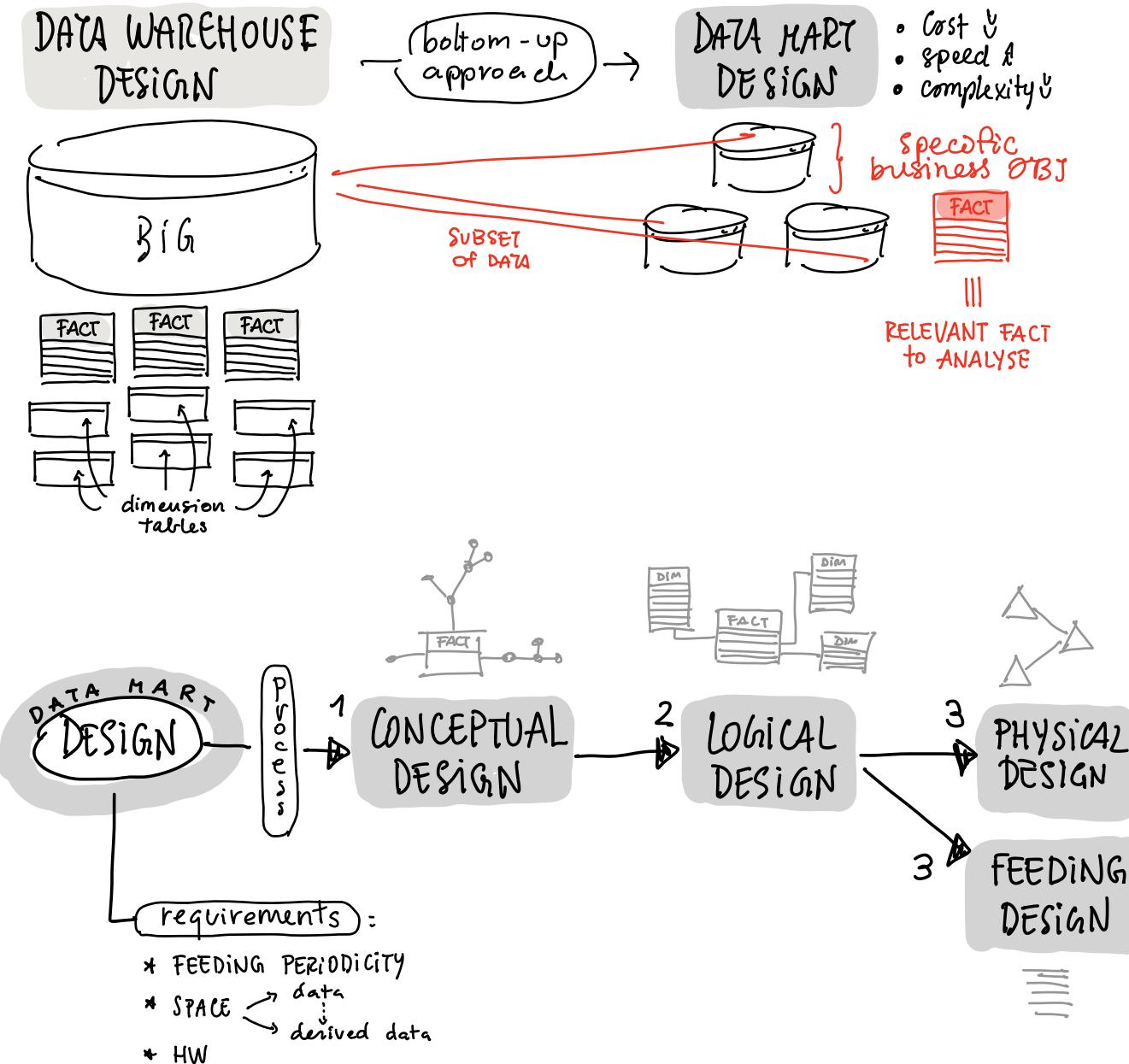
beatrice.occhiena

beatrice.occhiena@live.it

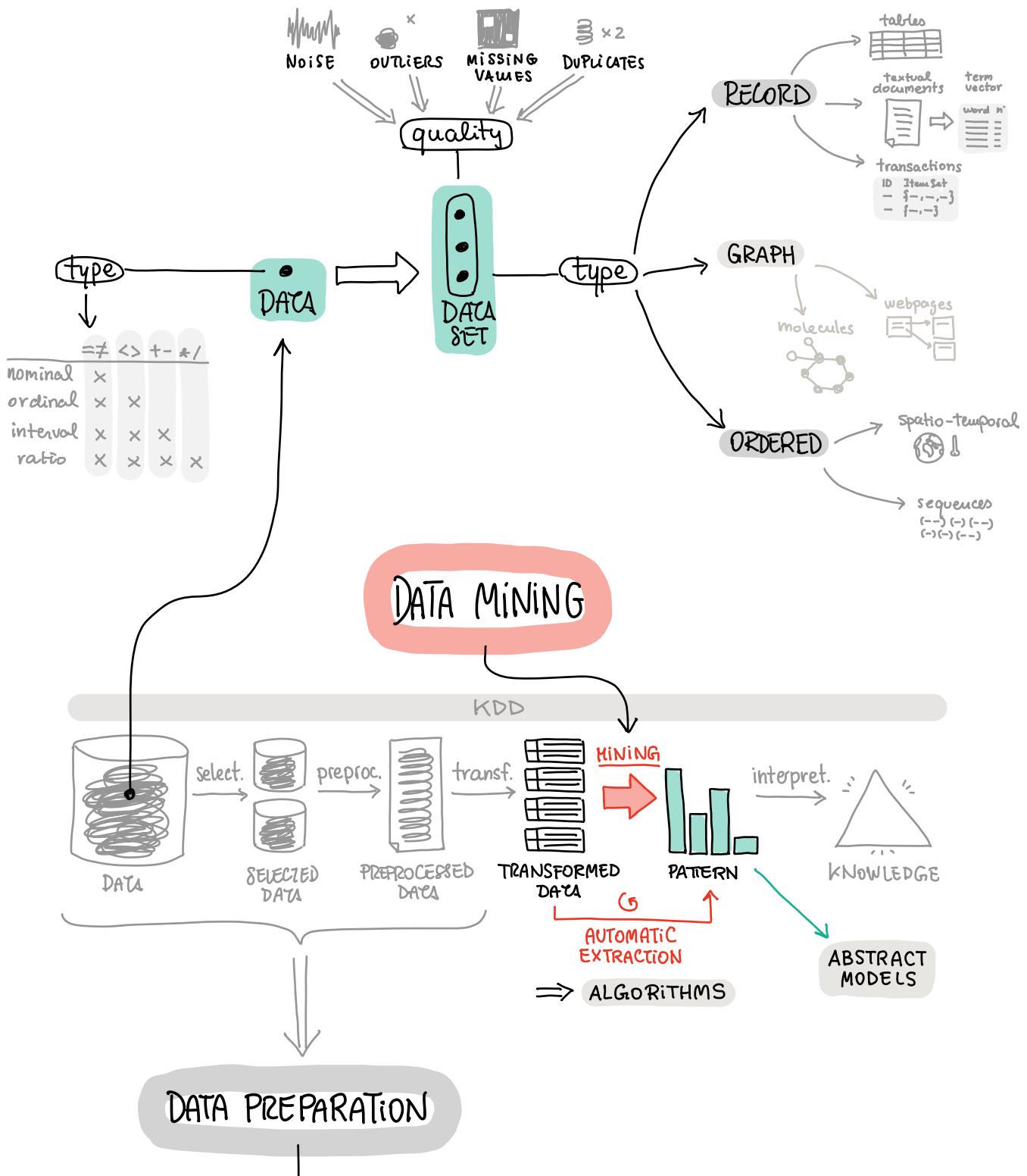
• PART 1: Introduction

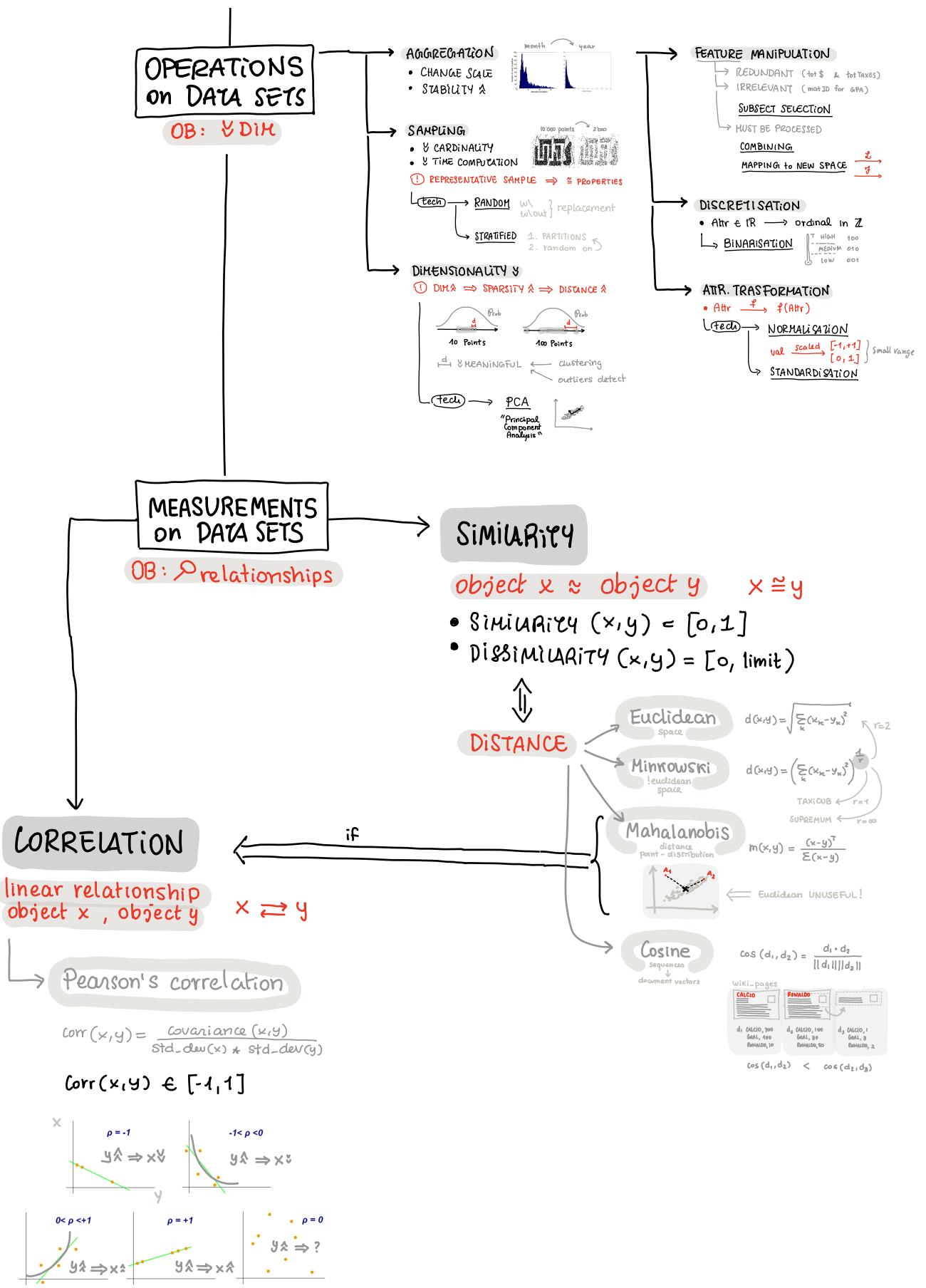






• PART 2 : Data Mining





DATA MINING APPROACHES



ASSOCIATION RULE LEARNING

Dependency

if ... → then ...



Forecasting
future behaviour

- Purchasing patterns
- Recommendations

CLASSIFICATION ANALYSIS

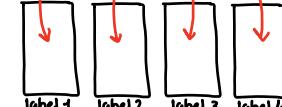
Sorting in \neq categories

based on features



new object

PREDEFINED CLASSES



DECISION BASED
ON PREV. TRAINING

Labeled responses
PREV. DATA SET

SUPERVISED
LEARNING

& CHECK MODEL'S
RESULTS

- ✉ → SPAM

- object recognition

CLUSTERING ANALYSIS

CREATE GROUPS
with relevant info
INHERENT in the set



NO LABELS!

GROUP BY
SIMILARITY

GROUP → $\approx \text{obj} \in G$
 $\approx \text{obj} \notin G$



UNSUPERVISED
LEARNING

- OUTLIERS DETECTION
- GEOGRAPHIC AREAS ID.



- UNDERSTANDING RELATIONSHIPS

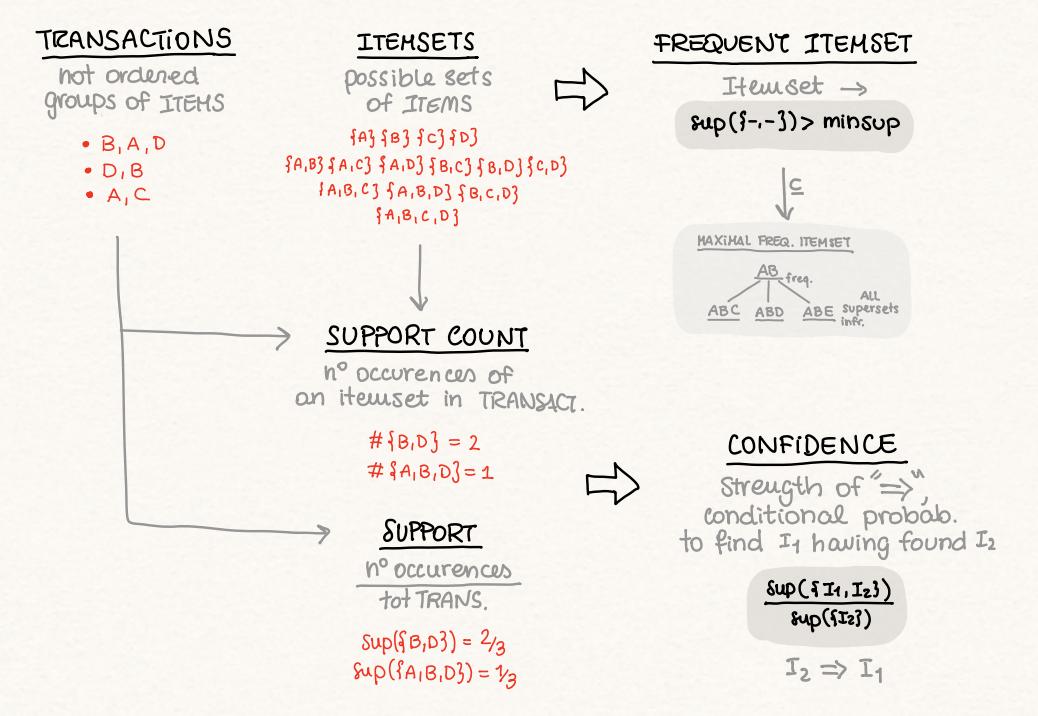
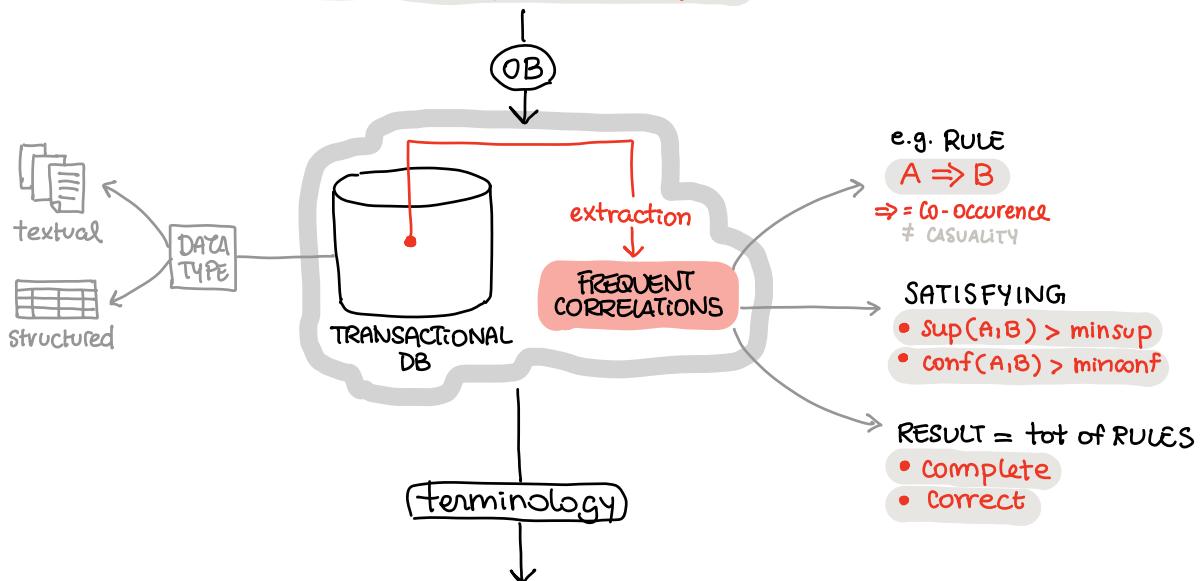


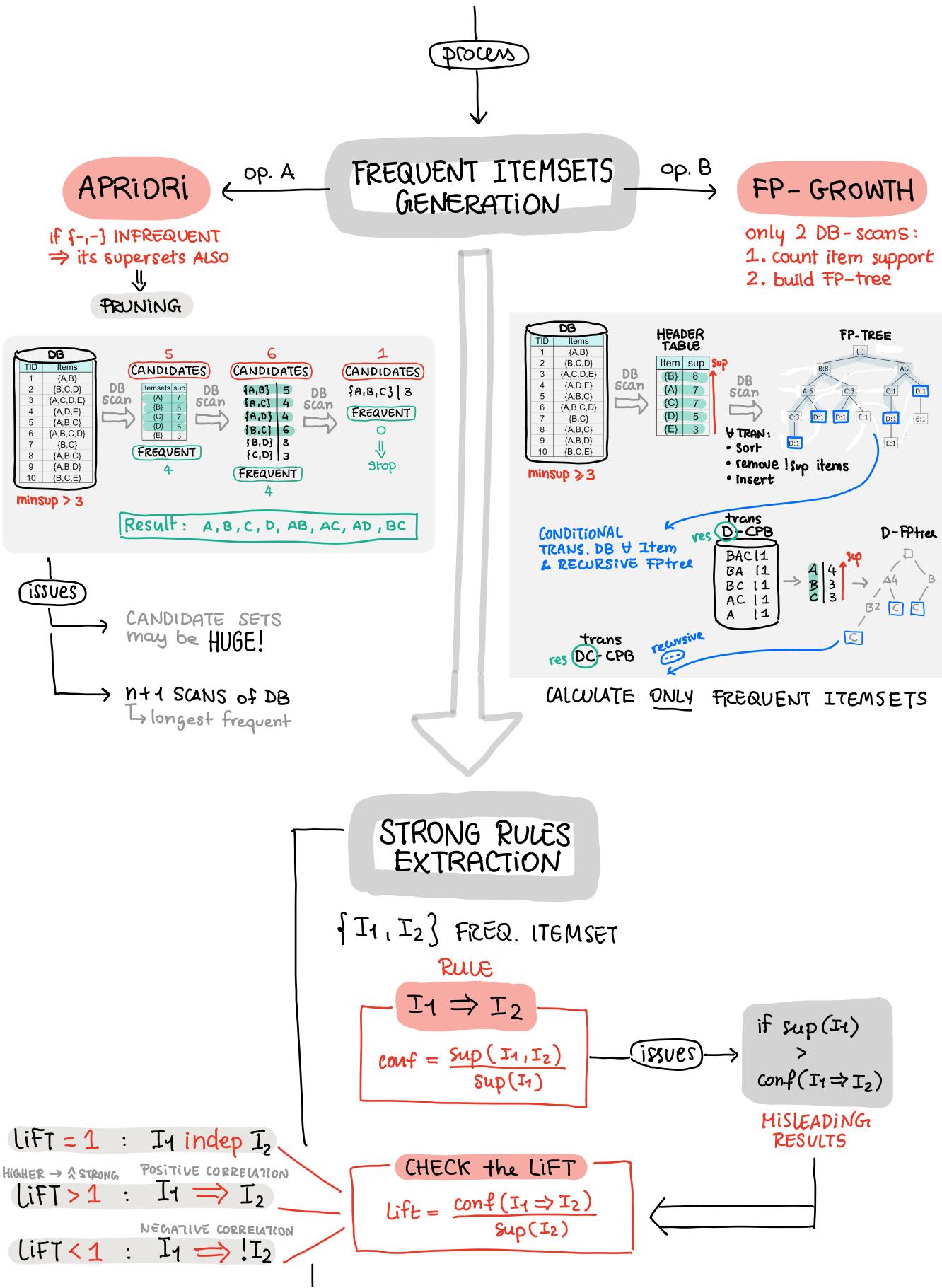
- SUMMARISATION

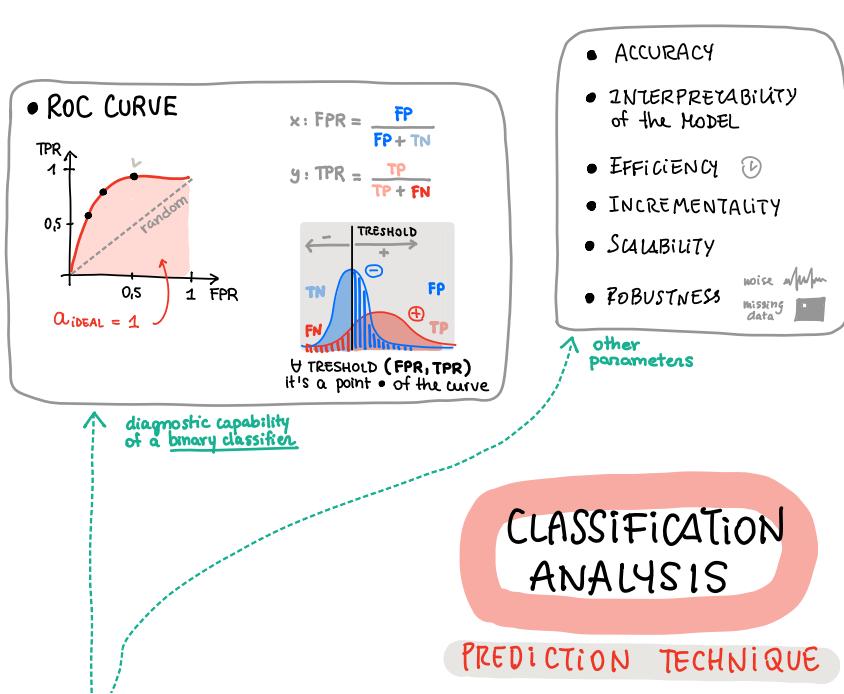
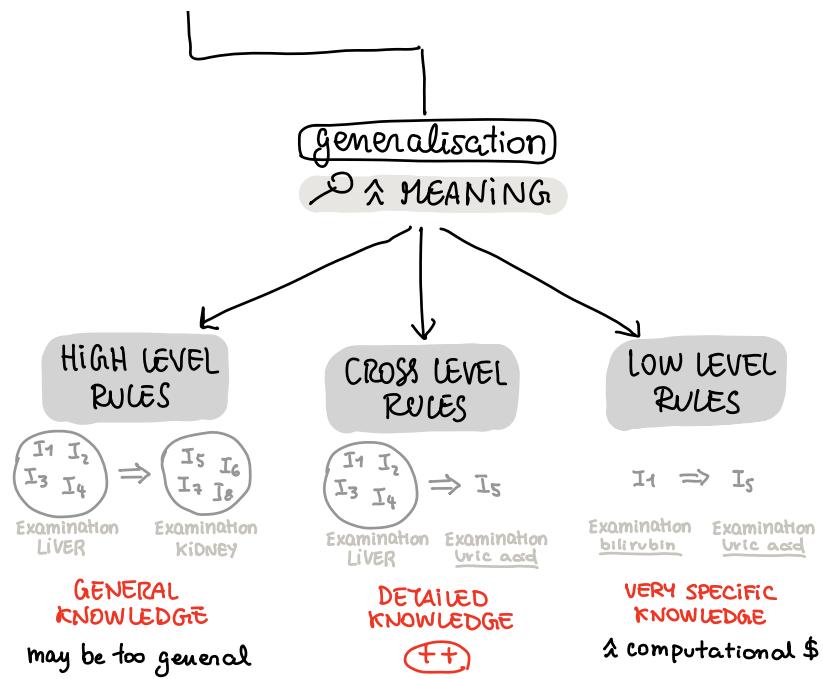


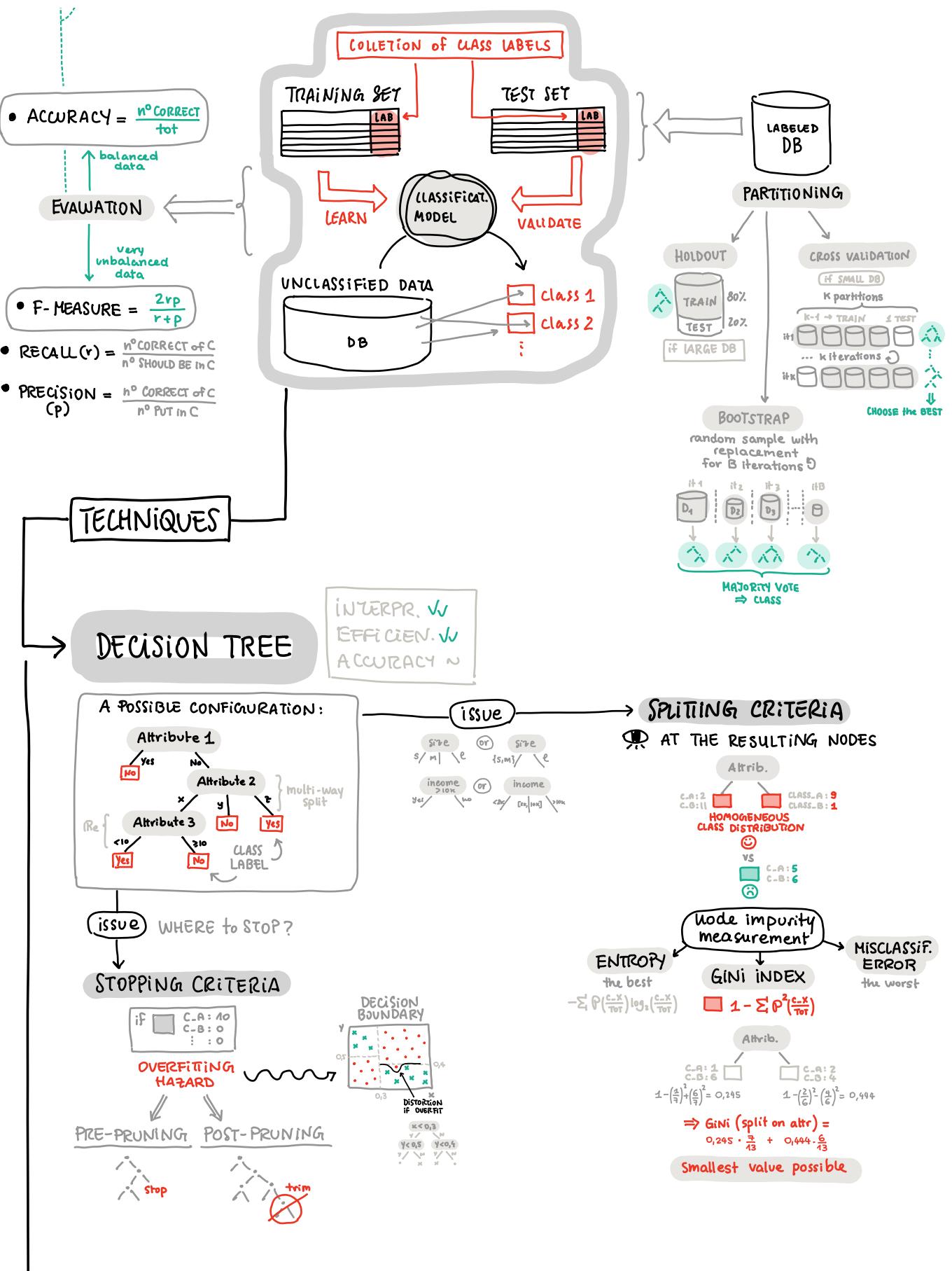
ASSOCIATION RULE LEARNING

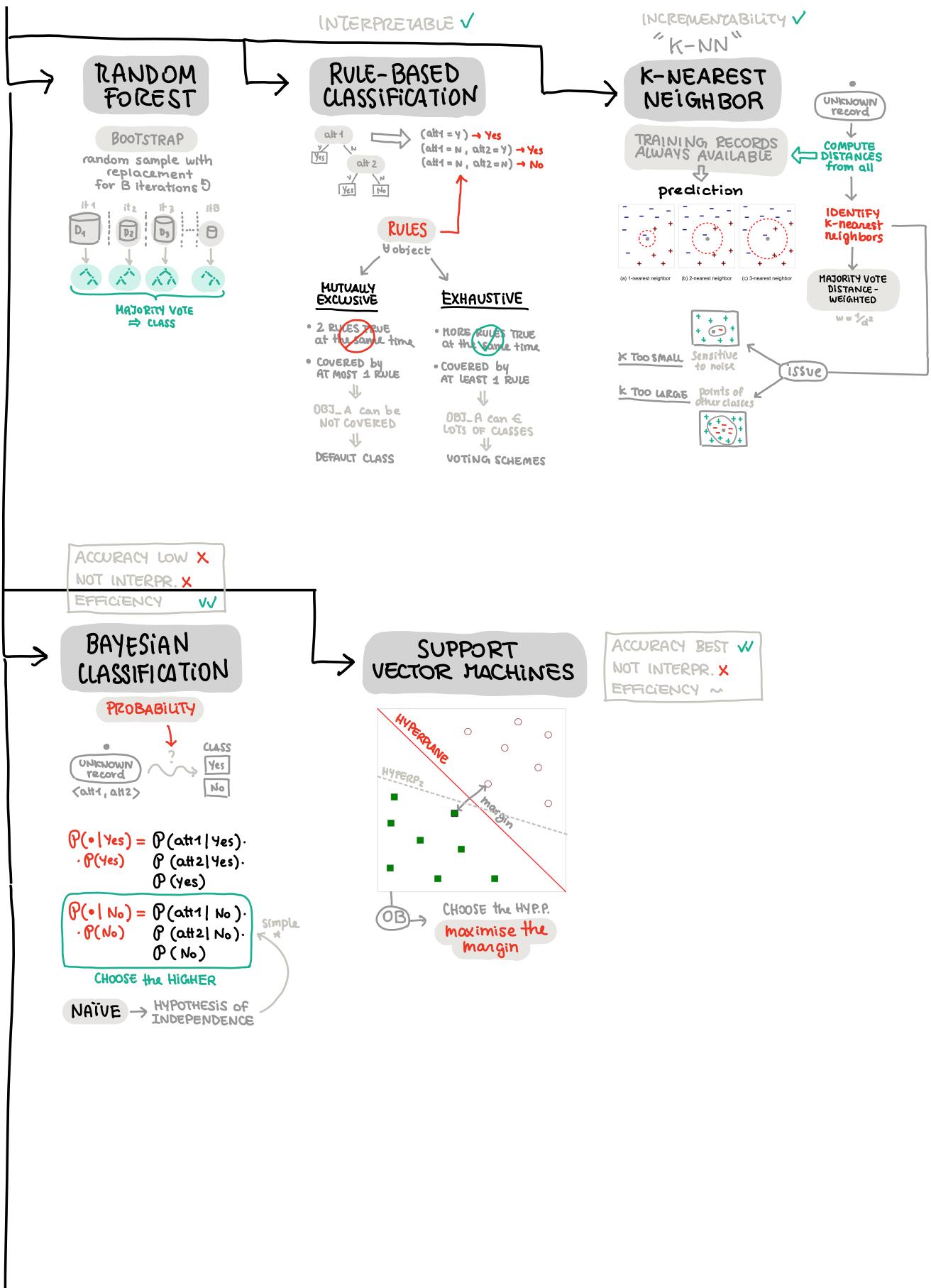
EXPLORATORY TECHNIQUE







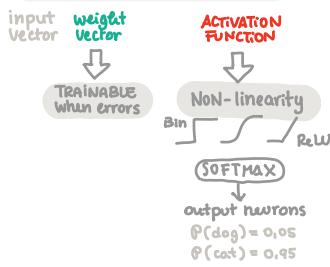
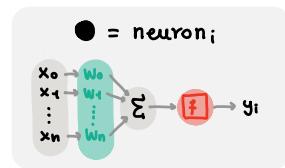
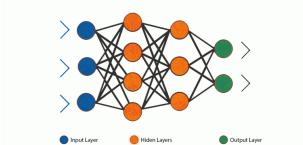




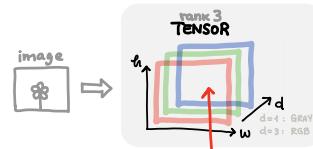
ARTIFICIAL NEURAL NETWORK

ACCURACY BEST ✓
NOT INTERPR. X
EFFICIENCY LOW X

FEED FORWARD NEURAL NETWORK "FFNN"

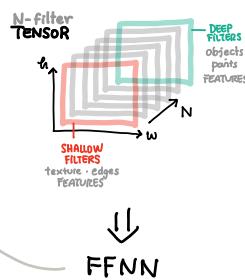


CONVENTIONAL NEURAL NETWORK "CNN"



AUTOMATICALLY EXTRACT FEATURES

100-1M Sliding filters
TRAINABLE WEIGHTS

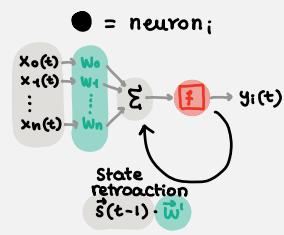


FFNN

RECURRENT NEURAL NETWORK "RNN"

PROCESS SERIANTIAL DATA

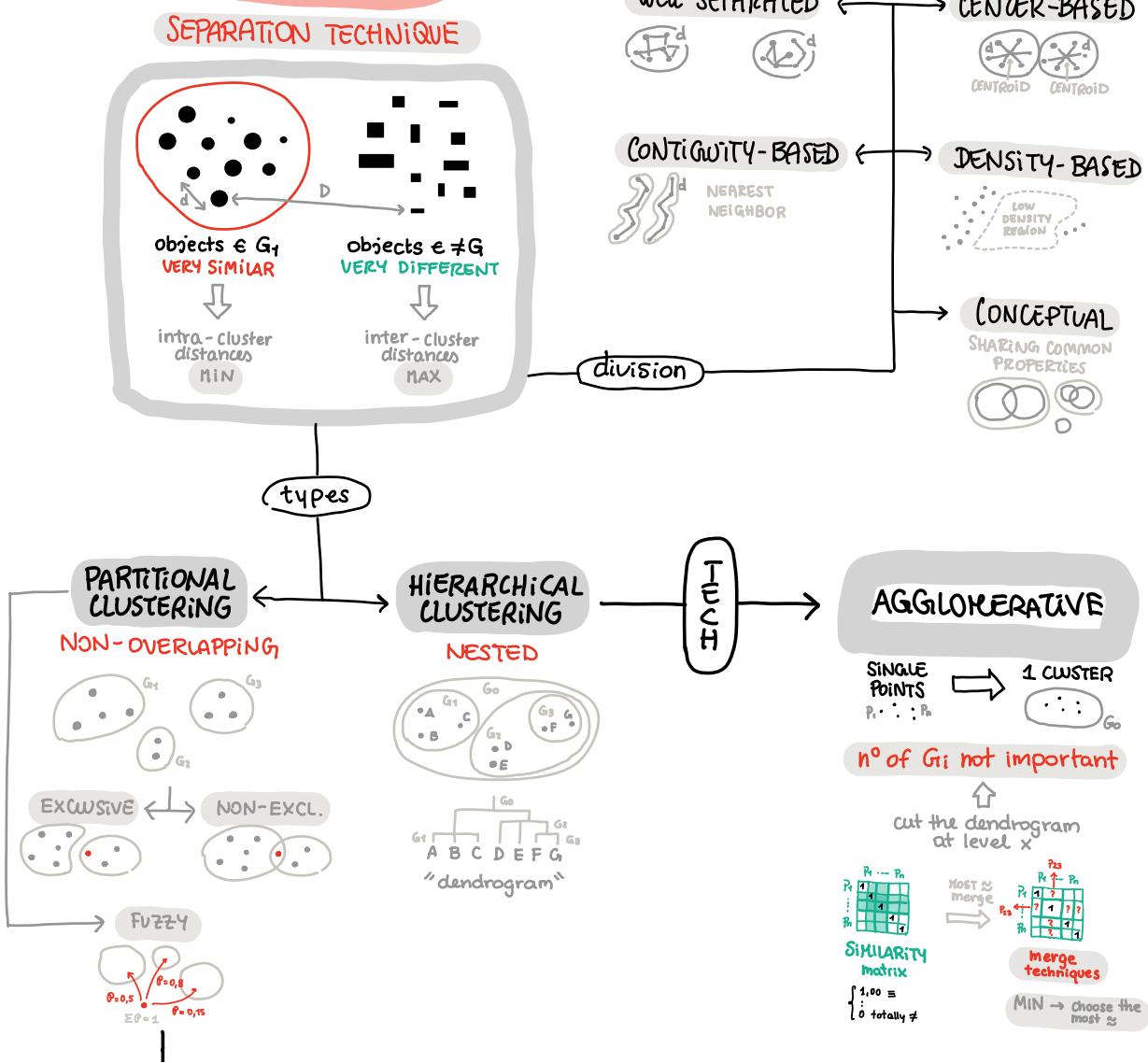
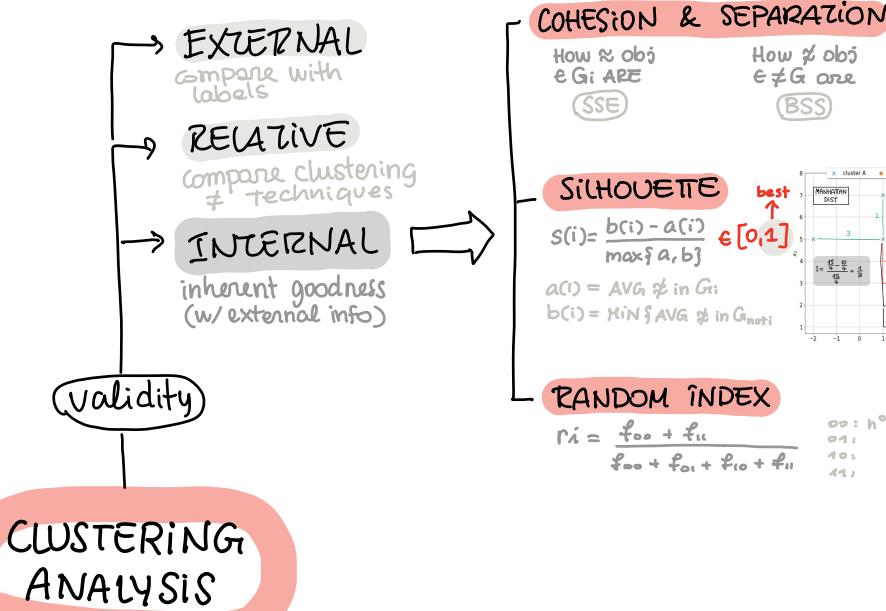
$\leftarrow \dots \rightarrow$ meaning



BACKPROPAGATION THROUGH TIME

LSTM

"Long Short Term Memories"

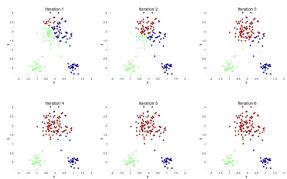
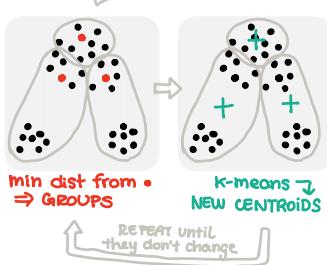
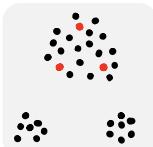


TECH

K-MEANS

CENTER BASED

INITIAL k^{known} CENTROIDS RANDOMLY CHOSEN



ISSUE

random init centroids

$\Rightarrow \neq$ RESULTS

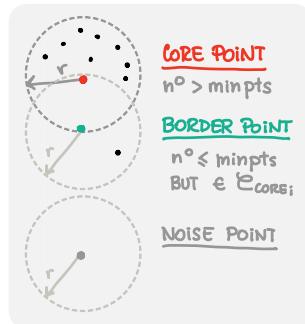
Problems with
↓

- \neq sizes
- \neq density
- Not globular shapes
- outliers

DB-SCAN

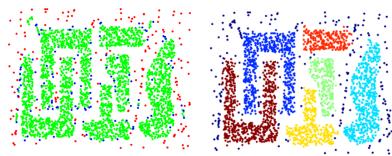
DENSITY BASED

n^{th} of points in \mathcal{E} ($r=eps$)



① ELIMINATE all noise pts

- ✓ RESISTANT to NOISE
- ✓ GOOD for \neq SHAPES



Problems with
↓

- Varying densities

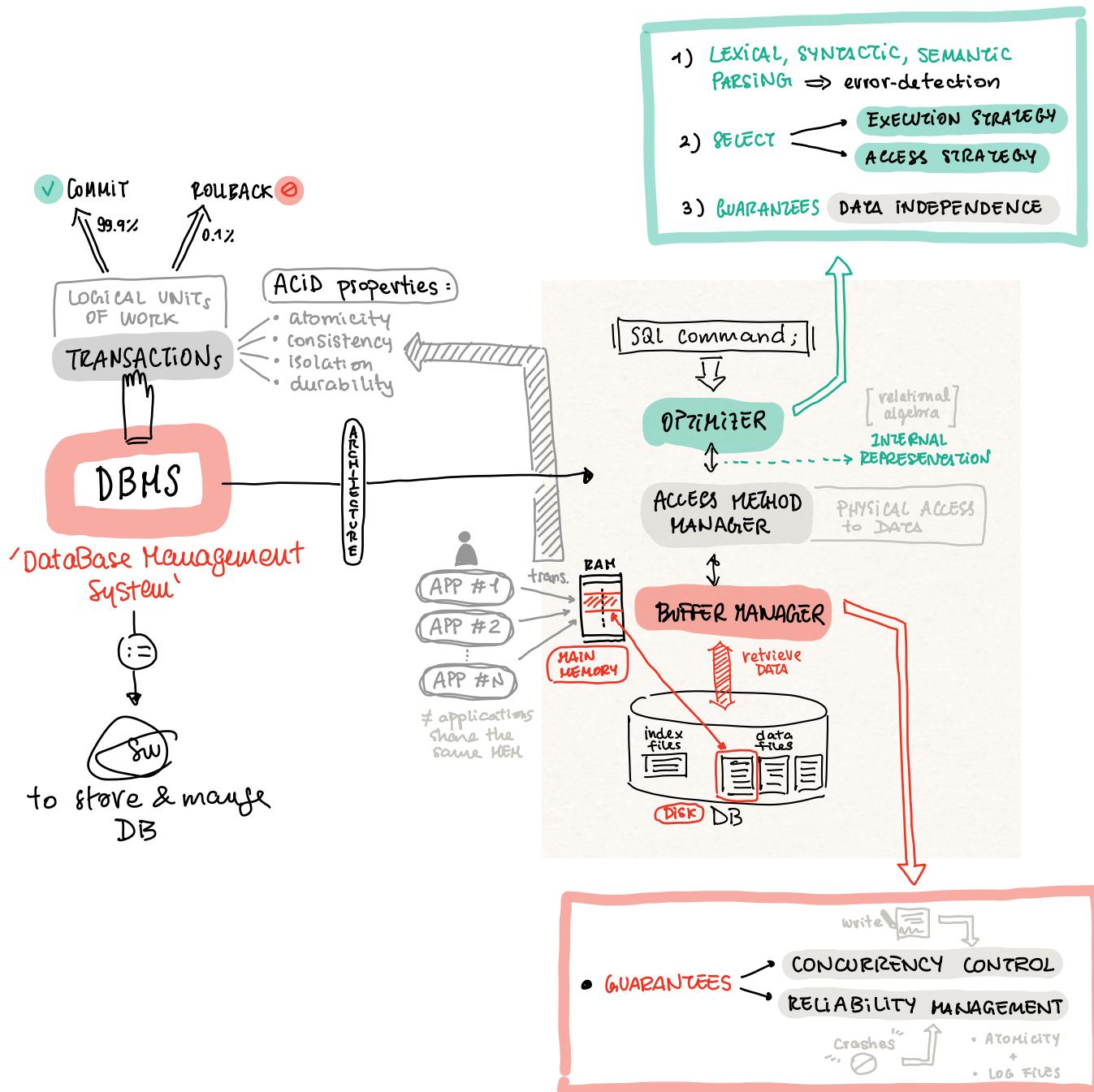


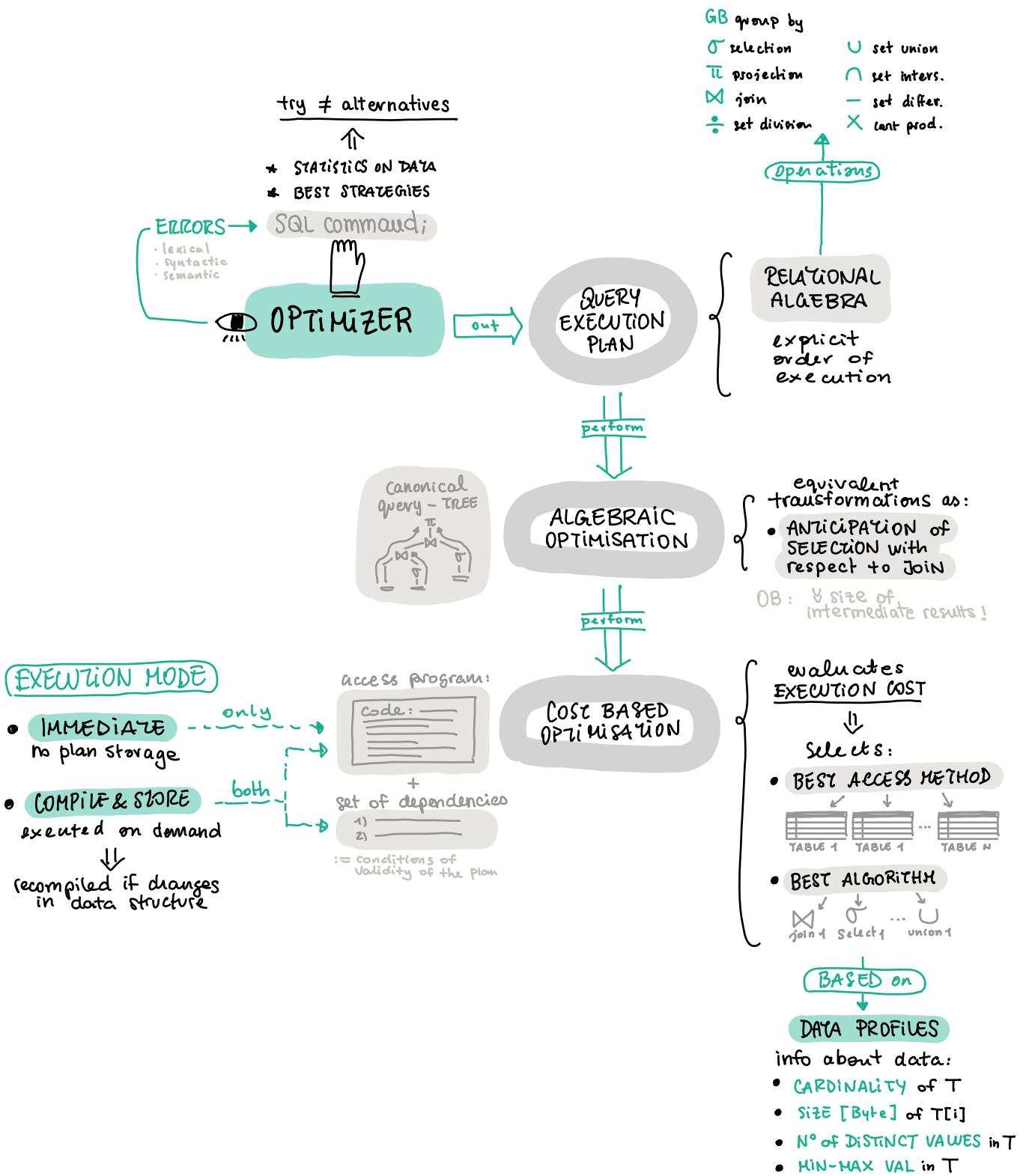
	11	12	13	14	15
11	1.00	0.90	0.10	0.65	0.20
12	0.90	1.00	0.70	0.60	0.50
13	1.00	0.70	0.40	0.30	
14	0.60	0.40	1.00	0.60	
15	0.50	0.30	0.60	1.00	
11	1.00	0.90	0.10	0.65	0.20
12	0.90	1.00	0.70	0.60	0.50
13	1.00	0.70	0.40	0.30	
14	0.65	0.40	1.00	0.60	
15	0.50	0.30	0.60	1.00	

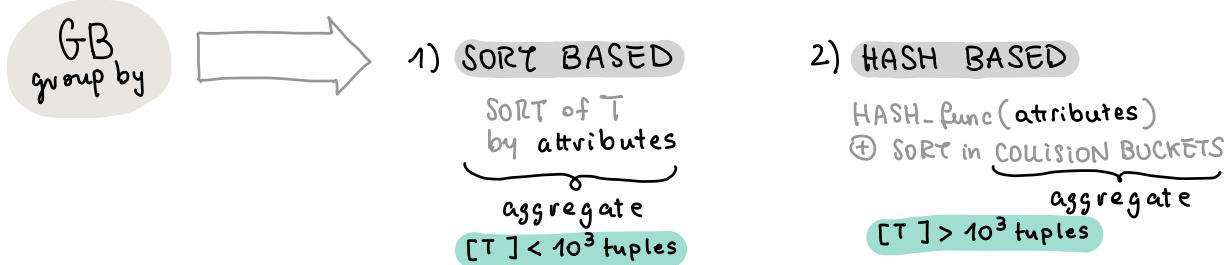
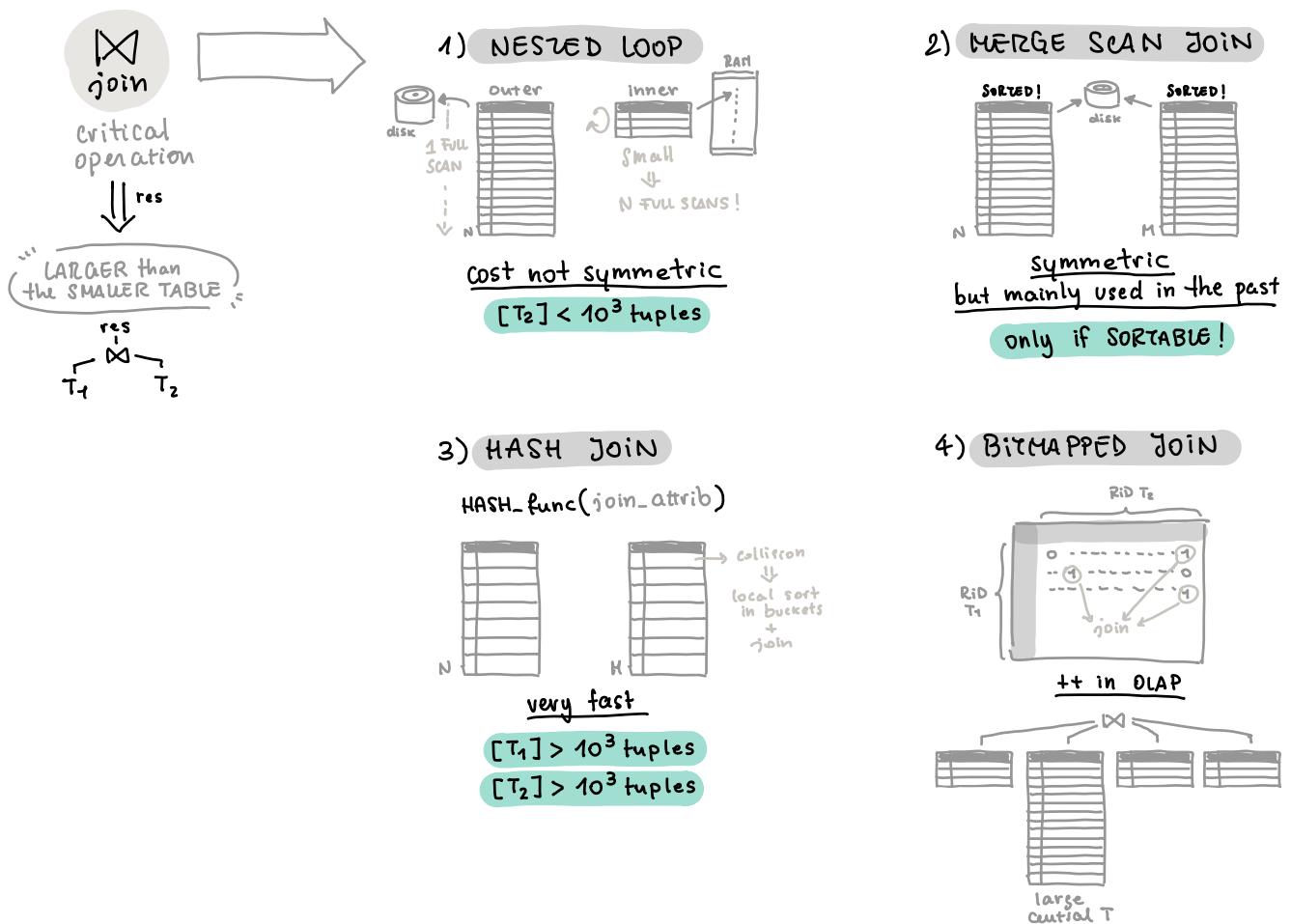
MAX → opposite

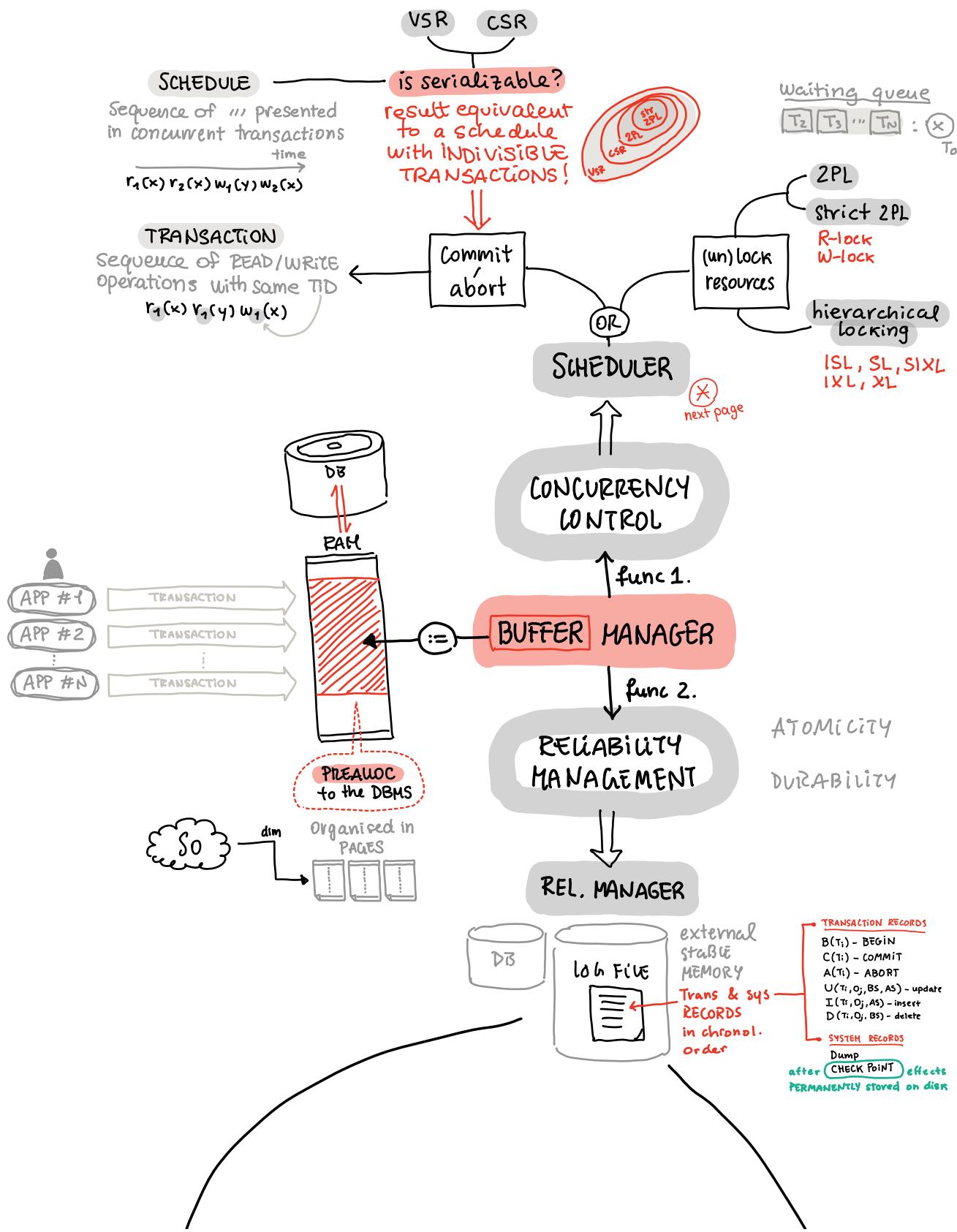
GROUP AVERAGE → the best

• PART 3 : DBMS









4

WRITING PROTOCOLS

OB: allow recovery if FAILURE



Before State 'BS' of DATA

FAILURE = T_i aborted but results in DB

allows UNDO



After state 'AS' of DATA

FAILURE = T_i commits but results ! in DB

allows REDO

for distributed transactions

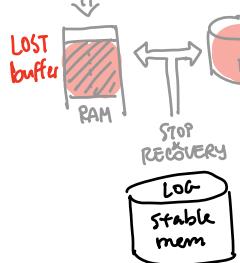


FAILURE

RECOVERY MANAGEMENT

SYSTEM

- power off
- sw problems

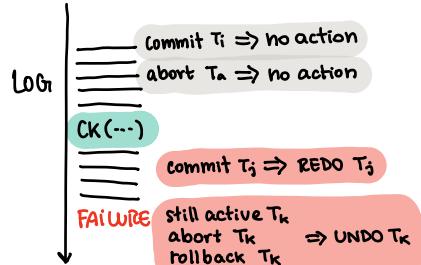


MEDIA

- disk damaged

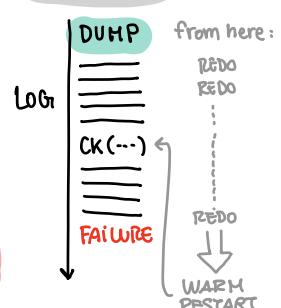
SYSTEM

WARM RESTART



MEDIA

COLD RESTART

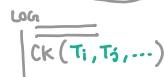


CHECK POINT

CK (list of active T_i)

WRITES on DISK the result of committed transactions

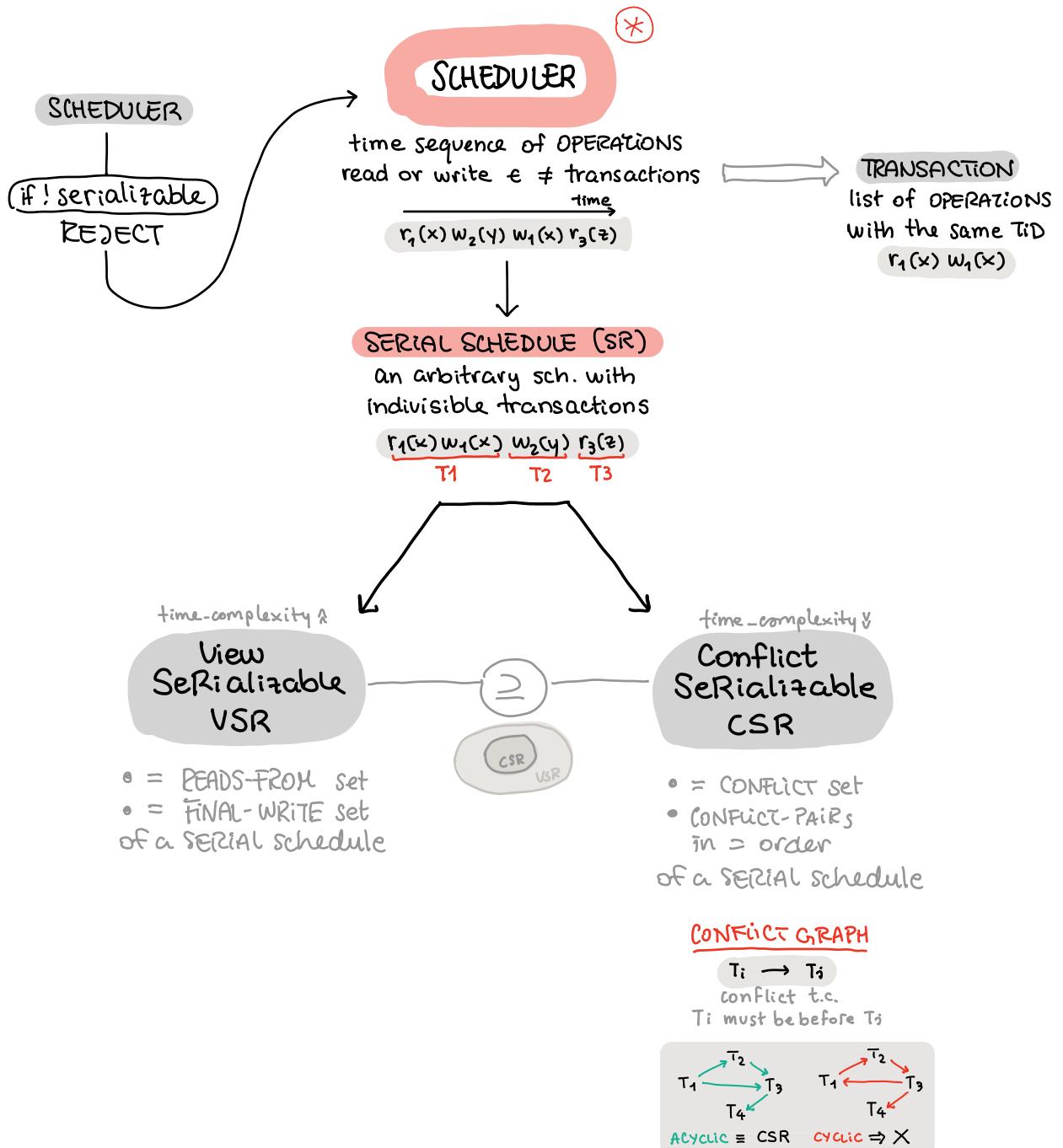
RECORDS active transactions



DUMP

complete COPY of the DB in Stable mem when the system is offline





• PART 4: Non-relational DB

